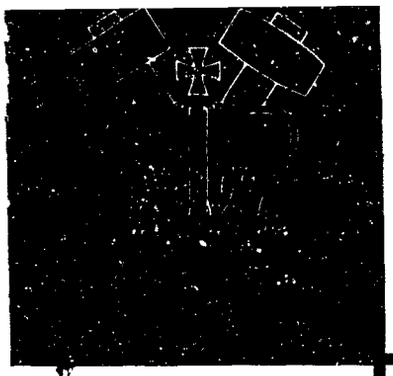
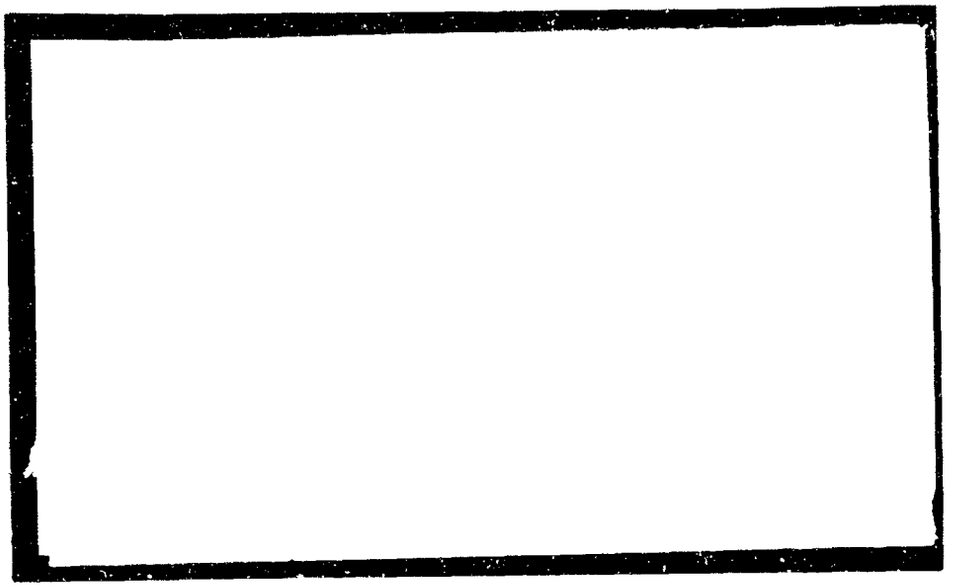


STATISTICAL ENGINEERING GROUP



COLUMBIA UNIVERSITY

SCHOOL OF ENGINEERING

NEW YORK 27 NEW YORK

Statistical Engineering Group
Columbia University
New York, New York

MARKOVIAN SEQUENTIAL CONTROL PROCESSES--

DENUMERABLE STATE SPACE

by

Cyrus Derman

Technical Report Number 23
August 1, 1964

This research was supported by the
Office of Naval Research under
Contract NONR - 266(55)

MARKOVIAN SEQUENTIAL CONTROL PROCESSES--
DENUMERABLE STATE SPACE [1]

Cyrus Derman

Columbia University

1. Introduction

As in [4], [5], [6] we are concerned with a dynamic system which is observed periodically and classified into one of a number of possible states. After each observation one of a possible number of decisions is made. The decisions determine the chance laws of the system. Previously, our considerations were confined to finite state spaces; here, we allow the number of possible states to be infinite.

Let I denote the state space of the system. Throughout, we shall assume I to be denumerable, though with suitable modifications our theorem below remains

- -

[1] Work sponsored by the Office of Naval Research under contract Nonr 266(55).

valid for a general state space. Whenever the system is in state i ($i \in I$) there are K_i possible decisions. Denoting by $\{Y_t\}$, $t=0,1,\ldots$, the sequence of states and by $\{\Delta_t\}$, $t=0,1,\ldots$, the sequence of decisions, we assume that

$$(M) \qquad P\{Y_{t+1} = j \mid s_{t-1}, Y_t = i, \Delta_t = k\} = q_{ij}(k)$$

for $k=1,\ldots,K_i$; $i,j \in I$; $t=0,1,\ldots$ where, for each t, s_t denotes the history of states and decisions (i.e., $s_t = \{Y_0 = y_0, \Delta_0 = d_0, \ldots, Y_t = y_t, \Delta_t = d_t\}$) and the $q_{ij}(k)$'s are non-negative numbers such that

$$\sum_{j \in I} q_{ij}(k) = 1 \quad , \qquad k=1,\ldots,K_i; \quad i \in I.$$

Roughly speaking, a rule R for sequentially controlling the process is a well-defined procedure which specifies the decision to be made at each point in time as a function of the history of the system. More precisely, we say R is a set of non-negative functions $\{D_k(s_{t-1}, y_t)\}$

where for each t (t=0,1,...) the domains of definition are the possible values of s_{t-1}, y_t and k and such that

$$\sum_k D_k(.,.) = 1 \quad ;$$

we define

$$P\left\{ \Delta_t = k \mid s_{t-1}, Y_t = y_t \right\} = D_k(s_{t-1}, y_t)$$

for all $k=1,\ldots,K_{y_t}$, s_{t-1}, y_t, and t=0,1,... . That is, we allow decisions to be made by a random mechanism, the mechanism used to depend on the history of the system.

We denote by $\mathcal{R}$, the class of all rules R. Once initial probabilities $P\{Y_0 = i\}$, $i \in I$, are given and a rule $R \in \mathcal{R}$ is specified, the sequences $\{Y_t\}$ and $\{Y_t, \Delta_t\}$, t=0,1,..., are stochastic processes. We shall call the process $\{Y_t, \Delta_t\}$ a <u>Markovian sequential control process</u>. It is not true that $\{Y_t\}$ or even $\{Y_t, \Delta_t\}$ will always be Markovian; whether they are or not will depend on the rule R. However, we use the term Markovian because of assumption M which imposes a kind of Markovian structure

on our processes. Such processes are a natural outgrowth of the dynamic programming point of view and the theory of Markov Chains. They were first discussed by Bellman (see e.g., [1] and [2] and also [4], [5], and [6] for other references.). Set

$$P_t(j,k \mid i,R) = P(Y_t = j,\ \Delta_t = k \mid Y_0 = i,R)$$

and let for any α, $0 < \alpha < 1$, $i \in I$

$$\phi(i,\alpha,R) = \sum_{t=0}^{\infty} \alpha^t \sum_{j,k} P_t(j,k \mid i,R)\, w_{jk}$$

where $\{w_{jk}\}$ are given numbers. $\phi(i,\alpha,R)$ can be thought of as the expected discounted cost over an infinite horizon of operating the system using rule R, given that i is the initial state and w_{jk} denotes the cost incurred whenever the system is in state j and decision k is made.

A question of concern is whether, for any given α and i, there exists a rule $R_0 \in \mathcal{R}$ such that

$$\psi(i,\alpha,R_0) = \min_{R\in\mathcal{R}} \psi(i,\alpha,R).$$

Conditions will be given which assure existence of such an optimal rule. It then follows that there is a non-randomized stationary rule which is optimal over $\mathcal{R}$. By a stationary rule we mean a rule such that

$$D_k(s_{t-1}, Y_t = i) = D_{ik} \quad ,$$

for every $t=0,1,\ldots$, $k=1,\ldots,K_i$, and $i\in I$. A non-randomized rule has its $D_k(.,.)$'s either zero or one. Thus a non-randomized stationary rule is such that there is one decision associated with each state and that decision is made each time the system is observed to be in that state.

2. Existence Theorem

Our result concerning the discounted cost criterion can be summarized as follows:

Theorem: If $K_i < \infty$ for each $i \in I$ and $\{w_{jk}\}$ is bounded, then for a given $\alpha (0 < \alpha < 1)$ there exists a non-randomized stationary rule R_0 such that

$$\psi(i,\alpha,R_0) = \min_{R \in \mathcal{R}} \psi(i,\alpha,R) \quad , \quad i \in I \quad .$$

Proof: The proof will fall into two parts: the first to show the existence of an optimal rule and the second to show that it can be taken to be a non-randomized stationary rule. The former, following the remarks of Karlin [7], involves showing that $\mathcal{R}$ is a compact space and $\psi(i,\alpha,R)$ is a continuous function over $\mathcal{R}$. The latter makes use of a device employed by Blackwell [3] in a similar proof for the case of a finite number of states.

If for a fixed n the collection of non-negative functions $\{D_k^{(n)}(.,.)\}$ is rule $R_n \in \mathcal{R}$, we say that $\lim_{n\to\infty} R_n = R \in \mathcal{R}$ if $\lim_{n\to\infty} D_k^{(n)}(.,.) = D_k(.,.)$ where $\{D_k(.,.)\}$ is the collection of non-negative functions constituting the rule R. In the following we arbitrarily set

$$P(Y_0 = i) = \beta_i \quad , \quad i \in I \quad ,$$

where $B_i \geq 0$, and $\sum_{i \in I} \beta_i = 1$.

First we have, as pointed out by Karlin [7],

Lemma 1. *If $K_i < \infty$ for each $i \in I$, then $\mathcal{R}$ is compact.*

Proof: For a fixed t, s_{t-1}, and y_t, the space consisting of the possible points

$$D^{(t)}(s_{t-1}, y_t) = \left\{D_1(s_{t-1}, y_t), \ldots, D_{K_{y_t}}(s_{t-1}, y_t)\right\}$$

is compact since $K_{y_t} < \infty$. By Tychonoff's theorem ([8] p. 260) the product space,

$$D^{(t)} = \prod_{s_{t-1}, y_t} D^{(t)}(s_{t-1}, y_t)\}$$

is compact; and again by the same theorem, the space

$$D = \prod_t D^{(t)}$$

is compact. However, D is the space $\mathfrak{R}$ of all rules. Hence, $\mathfrak{R}$ is compact.

Secondly, we shall prove

Lemma 2. Under the conditions of theorem 1, if $\lim_{n\to\infty} R_n = R \in \mathfrak{R}$ then for each t, t=0,1,...

$$\lim_{n\to\infty} \sum_{i\in I} \sum_{j,k} \beta_i \, P_t(j,k \mid i,R_n) \, w_{jk}$$

$$= \sum_{i\in I} \sum_{j,k} \beta_i \, P_t(j,k \mid i,R) \, w_{jk} \quad ;$$

consequently, $\sum_{i\in I} \beta_i \, \psi(i,\alpha,R)$ for fixed $\alpha (0 < \alpha < 1)$ is continuous over $\mathfrak{R}$.

Proof: We can write for any R_n, i, j, k and t

$$P_t(j,k \mid i,R_n) = \sum_{s_{t-1}} P(Y_t=j,\ \Delta_t=k,\ s_{t-1} \mid Y_0=i,\ R_n)$$

$$= \sum_{s_{t-1}} P(\Delta_t=k \mid Y_0=i,\ s_{t-1},\ Y_t=j,\ R_n)\ P(Y_t=j,\ s_{t-1} \mid Y_0=i,R_n)$$

$$= \sum_{s_{t-1}} D_k^{(n)}(s_{t-1},Y_t=j)\ P(Y_t=j \mid Y_0=i,\ s_{t-1},\ R_n)$$

$$\cdot P(s_{t-1} \mid Y_0=i,\ R_n)$$

$$= \sum_{s_{t-1}} D_k^{(n)}(s_{t-1},\ Y_t=j)\ q_{y_{t-1}j}\ (\Delta_{t-1})\ P(s_{t-1} \mid Y_0=i,\ R_n)$$

$$= \sum_{s_{t-1}} D_k^{(n)}(s_{t-1},\ Y_t=j)\ q_{y_{t-1}j}(\Delta_{t-1})\ D_{\Delta_{t-1}}^{(n)}(s_{t-2},\ y_{t-1})$$

$$\cdot\ q_{y_{t-2}y_{t-1}}(\Delta_{t-2}) \cdots q_{i_{y_1}}(\Delta_0) D_{\Delta_0}^{(n)}(Y_0=i)\ .$$

However, since for any $R \in \mathcal{R}$

$$\sum_{j,k} P_t(j,k \mid i,R) = 1 \quad .$$

and the $D_k^{(n)}(.,.)$'s converge, it follows from a theorem due to Scheffe [9] that

$$\lim_{n\to\infty} \sum_{j,k\in E} P_t(j,k \mid i,R_n) = \sum_{j,k\in E} P_t(j,k \mid i,R)$$

for any set E in the space of possible states and decisions. However, since $\{w_{jk}\}$ is bounded, the lemma follows using standard arguments.

We remark that since $\sum_{i\in I} \sum_{j,k} \beta_i P_t(j,k \mid i,R)\, w_{jk}$ is bounded as well as continuous, it follows that for fixed $\alpha (0 < \alpha < 1)$

$$\sum_{i\in I} \beta_i \, \psi(i,\alpha,R) = \sum_{t=0}^{\infty} \alpha^t \sum_{i\in I} \sum_{j,k} \beta_i \, P_t(j,k \mid i,R) \, w_{jk}$$

is also continuous over $\mathcal{R}$.

Combining lemmas 1 and the above remark we have

Lemma 3: Under conditions of theorem 1, for a given $\alpha(0 < \alpha < 1)$ there exists a rule $R^* \in \mathcal{R}$ such that

$$\psi(i,\alpha,R^*) = \min_{R\in\mathcal{R}} \psi(i,\alpha,R) \quad , \quad i\in I \quad .$$

Proof: From the well-known fact that a continuous function achieves its minimum over a compact space we have from lemma 1 and the remark after lemma 2 that there exists a rule R* such that

$$\sum_{i\in I} \beta_i \, \psi(i,\alpha,R^*) = \min_{R\in\mathcal{R}} \sum_{i\in I} \beta_i \, \psi(i,\alpha,R) \quad .$$

However, suppose that β_i's are chosen so that $\beta_i > 0$,

$i \in I$; then R* must be as asserted in the lemma. For otherwise we could construct a different rule which would provide a smaller values of $\sum_{i \in I} \beta_i \, \psi(i,\alpha,R)$.

We now proceed to the second part of the proof of the theorem; namely, to show that there exists a non-randomized stationary rule R_0 such that

$$\psi(i,\alpha,R_0) = \psi(i,\alpha,R^*) \quad , \quad i \in I \quad ,$$

where R* is as in lemma 3.

Following Blackwell [3], if D denotes the set of numbers $\{d_{ik}\}$, $d_{ik} \geq 0$, $\sum_k d_{ik} = 1$, $i \in I$, then let $R_1 = (D,R^*)$ denote the rule:

$$D_k(Y_0 \, , \, i) = d_{ik} \quad , \quad k=1,\ldots,K_i, \quad i \in I \quad ,$$

followed by use of the rule R* for the process

$\{Y'_{t-1} = Y_t \ , \ \Delta'_{t-1} = \Delta_t\}$, t=1,... . More generally, let $R_n = \{D,\ldots,D,\ R^*\}$ denote the rule:

$$D_k(s_{t-1},\ Y_t = i) \ = \ d_{ik},\ k=1,\ldots,K_i,\quad i\in I,\quad 0 \le t \le n \ ,$$

followed by use of the rule R* for the process $\{Y^{(n)}_{t-n} = Y_t \ ,$ $\Delta^{(n)}_{t-n} = \Delta_t\}$, t=n,... .

Let $\{d_{ik}\}$ be chosen so that, for each $i\in I$,

$$\sum_k d_{ik}\ w_{ik} + \alpha \sum_{j,k} q_{ij}(k)\ d_{ik}\ \psi(j,\alpha,R^*)$$

is minimized. Clearly, the minimizing values can be taken to be zero or one. From such a choice of $D = \{d_{ik}\}$, it is easily seen that, for n=1,2,...,

$$\psi(i,\alpha,R_n) = \psi(i,\alpha,R^*) \quad , \quad i\in I \ .$$

Moreover, $\lim_{n\to\infty} R_n = R_0$, the non-randomized stationary rule with $\{D_{ik}\} = D$. However, by lemma 2, $\psi(i,\alpha,R)$ is continuous over $\mathcal{R}$; hence

$$\psi(i,\alpha,R_0) = \lim_{n\to\infty} \psi(i,\alpha,R_n)$$

$$= \psi(i,\alpha,R^*) \quad , \quad i\in I \ .$$

This last equation establishes the theorem.

3. Counter-Example

There is no difficulty in providing an example in which the condition of finiteness of the K_i's is violated and the conclusion of the theorem does not hold.

The following example shows that the theorem may not hold if the boundedness condition on $\{w_{jk}\}$ is weakened. Let I consist of the states $0, 1_a, 1_b, 2_a, 2_b, \ldots$ and suppose there are two possible decisions

at states 1_a, 2_a,... and only one possible decision at states 0, 1_b, 2_b,... . Assume that

$$q_0(1) = 1, \quad q_{i_b i_b}(1) = 1 \quad , \qquad i=1,2,...$$

$$q_{i_a (i+1)_a}(1) = p \quad , \quad q_{i_a 0}(1) = 1-p \quad , \quad i=1,2,..., \quad 0 < p < 1 \quad ,$$

$$q_{i_a i_b}(2) = 1 \qquad i=1,2,...;$$

$$w_{01} = 0 \quad , \quad w_{i_b 1} = - \frac{(i-1)}{(\alpha p)^{i-1}} \quad , \qquad i=1,2,...$$

$$w_{i_a 1} = w_{i_a 2} = - \frac{1}{(2\alpha p)^{i-1}} \quad , \qquad i=1,2,...$$

Let $P(Y_0 = 1_a) = 1$. If R_n denotes the rule: Make decision 1 for all $t < n$; if $Y_n = (n+1)_a$ make decision 2 at $t = n$. Then, on computing ψ, we get

$$\psi(1_a,\alpha,R_n) = w_{1_a1} + \alpha p w_{2_a1} + \ldots + (\alpha p)^n w_{(n+1)_a2} +$$

$$(\alpha p)^n \frac{\alpha}{1-\alpha} \; w_{(n+1)_b1}$$

$$= -\left[\left(1 + \frac{1}{2} + \ldots + \frac{1}{2^n}\right) + \frac{\alpha}{1-\alpha}\, n \right] .$$

Thus $\lim_{n\to\infty} \psi(1_a,\alpha,R_n) = -\infty$. However, every $R\in\mathcal{R}$ will clearly yield a finite value for $\psi(1_a,\alpha,R)$. Thus no optimal rule exists.

4. Remarks

Of interest are conditions under which the assertion of the theorem holds when ψ is replaced by

$$Q_R(i) = \lim_{T\to\infty} \sup \frac{1}{T} \sum_{t=1}^{T} \sum_{j,k} P_t(j,k \mid i,R)\, w_{jk} \quad .$$

When the limit exists this is usually referred to as the average cost per unit time. It was shown in [4] that the theorem holds when I is finite. However, a proof in the denumerable case has not been given and it is not entirely clear that it is true, not withstanding the usual intuitive arguments.

For I finite, Blackwell [3] obtained a stronger result. He showed there exists a non-randomized stationary rule R_0 such that

$$\psi(i,\alpha,R_0) = \min_{R\in\mathfrak{R}^*} \psi(i,\alpha,R) \quad , \quad i\in I$$

for every α near enough but less than one. $\mathfrak{R}^*$ is the class of all rules whose decisions at time t depend only on the state Y_t and t. However, from the above result it is clear that $\mathfrak{R}^*$ can be replaced by $\mathfrak{R}$. A counter-example[1] appears in the doctoral thesis of Ashok Maitra (Department of Statistics, University of California, Berkeley) indicating that the result does not extend to the denumerable case.

- -

[1] Communicated to me by David Blackwell.

References

[1] Bellman, Richard (1957). Dynamic Programming. Princeton University Press.

[2] Bellman, Richard (1957). A Markovian decision process. J. Math. Mech. 6 679-684.

[3] Blackwell, David (1962). Discrete dynamic programming. Ann. Math. Statist. 33 719-726.

[4] Derman, Cyrus (1962). On sequential decisions and Markov chains. Management Sci. 9 16-24.

[5] Derman, Cyrus (1963). Stable sequential rules and Markov chains. J. Math. Analysis and Applications 6 257-265.

[6] Derman, Cyrus (1964). On sequential control processes. Ann. Math. Statist. 35 341-349.

[7] Karlin, Samuel (1955). Structure of dynamic programming. Naval Research Logistics Quarterly 2 285-294.

[8] McShane, E.J. and Botts, T. (1959). Real Analysis Von Nostrand.

[9] Sheffé, Henry (1947). A useful convergence theorem for probability distributions. Ann. Math. Stat. 18 434-438.